

EXTENSÃO BIVARIADA DO ÍNDICE DE CONFIABILIDADE UNIVARIADO PARA AVALIAÇÃO DA ESTABILIDADE FENOTÍPICA

Bivariate extension of univariate reliability index for evaluating phenotypic stability

Suzankelly Cunha Arruda de Abreu¹, Daniel Furtado Ferreira²,
Fábio de Lima Gurgel³, Ângela de Fátima Barbosa Abreu⁴

RESUMO

Com o presente trabalho, objetiva-se realizar a derivação teórica da extensão bivariada dos métodos de Annicchiarico (1992) e Annicchiarico et al. (1995) para estudar a estabilidade fenotípica. A partir dos ensaios com ρ genótipos em q ambientes e mensurações de duas variáveis, cada genótipo teve seu valor padronizado com relação a cada variável $k = 1, 2$. Essa padronização foi realizada em função da média do ambiente, da seguinte forma: $W_{ijk} = Y_{ijk} / \bar{Y}_{.jk} \times 100$; em que W_{ijk} representa o valor padronizado do genótipo i , no ambiente j para a variável k ; Y_{ijk} representa a média observada do genótipo i , no ambiente j para a variável k e $\bar{Y}_{.jk}$, a média de todos os genótipos para o ambiente j e variável k . Com os valores padronizados foram estimados o vetor média e a matriz de variância e covariância de cada genótipo. Foi obtida a derivação teórica da extensão bivariada do índice de risco (I_i) de Annicchiarico com sucesso e foi proposto um segundo índice de risco baseado nas probabilidades bivariada (Prb_i); os dois índices apresentaram grande concordância nos resultados obtidos em um exemplo ilustrativo com genótipos de melões.

Termos para indexação: Índice de risco, distribuição normal bivariada, componentes principais.

ABSTRACT

The objective of this work was to obtain the theoretical derivation of the bivariate extension to the methods proposed by Annicchiarico (1992) and Annicchiarico et al. (1995) for studying phenotypic stability. Considering assays with ρ genotypes in q environments and two variates, every genotype had the response of each variate ($k = 1, 2$) standardized. This standardization has been made using the environment means as follows: $W_{ijk} = Y_{ijk} / \bar{Y}_{.jk} \times 100$; where W_{ijk} represents the i th genotype standard value in the j th environment for the k th variate; Y_{ijk} represents the observed mean of the i th genotype, in j th environment for the k th variate and $\bar{Y}_{.jk}$ the overall genotypes means for j th environment to k th variate. Considering the standardized values, the genotypes mean vector and covariance matrix were estimated. The theoretical derivation of the bivariate reliability index (I_i) of Annicchiarico was worked out and a second risk index was proposed based on the bivariate probabilities (Prb_i); the two indexes proposed has shown good agreement to the results in an example with melon genotypes.

Index terms: Reliability index, bivariate normal distribution; principal components.

(Recebido para publicação em 7 de novembro de 2003 e aprovado em 27 de janeiro de 2004)

INTRODUÇÃO

Os estudos da estabilidade fenotípica têm sido objetos de muitos estudos nas últimas décadas, por estarem envolvidos com uma das mais importantes fases dos programas de melhoramento de plantas. A filosofia básica desses estudos consiste em relacionar a produtividade média do ambiente, expressa por um índice ambiental, com as respostas dos genótipos. Implícita-

mente, um dos objetivos de muitos programas de melhoramento é desenvolver e identificar genótipos que combinem elasticidade, em ambientes pobres, com responsividade nos ambientes de alta tecnologia (Becker e Leon, 1988; Crossa, 1990).

Dentre as metodologias mais simples, destaca-se o uso do índice de risco de cada genótipo proposto por Annicchiarico (1992) e Annicchiarico et al. (1995).

1. Aluna do 9º período do curso de Administração da Universidade Federal de Lavras/UFLA – Caixa Postal 3037 – 37200-000 – Lavras, MG, bolsista CNPq, suzankellys@bol.com.br

2. Professor Adjunto IV do Departamento de Ciências Exatas da UFLA, bolsista CNPq, danielff@ufla.br

3. Aluno de doutorado do Curso de Agronomia – Área de Concentração em Genética e Melhoramento de Plantas – Departamento de Biologia da UFLA, bolsista CNPq, flgurgel@uol.com.br

4. Pesquisadora da Embrapa lotada no Departamento de Biologia da UFLA, afbabreu@ufla.br

O índice de risco refere-se a um parâmetro de avaliação da estabilidade fenotípica e é estimado nessa metodologia. Esses autores propõem que se considere um coeficiente de confiança de 75% e que as médias genotípicas sejam expressas em termos da porcentagem dos valores médios dos ambientes. Posteriormente, estimam-se, com os dados dessa forma padronizados, a média e o desvio padrão de cada genótipo em relação aos ambientes. Utilizando-se esses valores e adotando-se a distribuição normal como modelo, os índices de risco são estimados, os quais fornecem a idéia, fixada uma probabilidade de confiança, da proporção mínima que um determinado genótipo supera a média dos ambientes estudados. Quanto maior o valor do índice de risco, mais estável será considerado o genótipo.

Destacam-se também nos estudos de estabilidade fenotípica as metodologias multivariadas. Dentre elas, pode-se citar o método Multivariado AMMI (Additive Multiplicative Models Interaction) (GAUCH JUNIOR e ZOBEL, 1988; CROSSA, 1990; YAU, 1995) que vem sendo largamente utilizado. Esse método baseia-se inicialmente na estimação de efeitos aditivos para os genótipos e para os ambientes, pelo procedimento usual da análise de variância (ANAVA). Em seguida, devem ser estimados os efeitos da interação genótipos x ambientes, considerados multiplicativos, utilizando a análise dos componentes principais (ACP) (JOHNSON e WICHERN, 1998), de acordo com Crossa (1990). O método AMMI não é usado como um método genuinamente multivariado, sendo aplicado em várias variáveis fenotípicas. Considera-se, no entanto, como variável ora os efeitos de interação para cada ambiente e ora os efeitos de interação para cada genótipo.

Finalmente, é conveniente salientar que o tratamento multivariado e a coleta de informações de vários caracteres nos ensaios das etapas finais do melhoramento vêm sendo uma prática constante (BECKER e LEON, 1988; ACCIARESI e CHIDICHIMO, 1999). Nenhuma menção, entretanto, foi encontrada na literatura com relação à generalização para o caso multivariado da metodologia do índice de risco (reliability index).

Em função do exposto, com o presente trabalho objetiva-se de obter a derivação teórica da extensão bivariada dos métodos de Annicchiarico e propor um segundo índice de risco baseado no modelo normal bivariado padrão dos dados padronizados. Um exemplo ilustrativo foi apresentado.

METODOLOGIA

Inicialmente foi postulado que a resposta média do *i*-ésimo genótipo, no *j*-ésimo ambiente, na *k*-ésima variável é dada por Y_{ijk} , sendo, $i = 1, 2, \dots, p$; $j = 1, 2, \dots, q$; $k = 1, 2$. A generalização multivariada do índice de risco de cada genótipo proposto por Annicchiarico (1992) e Annicchiarico et al. (1995) se deu da seguinte forma: inicialmente, cada genótipo teve seu valor padronizado com relação a cada variável $k = 1, 2$. Essa padronização foi realizada com relação à média do ambiente, da seguinte forma:

$$W_{ijk} = \frac{Y_{ijk}}{\bar{Y}_{\bullet jk}} \times 100 \quad (2.1)$$

em que W_{ijk} representa o valor padronizado do genótipo *i*, no ambiente *j* para a variável *k*; $\bar{Y}_{\bullet jk}$ é o valor médio do ambiente *j* na *k*-ésima variável em relação a todos os genótipos.

De posse dos valores padronizados na equação (2.1), foram estimados o vetor de médias ($\bar{W}_{i\bullet}$) e a matriz de variâncias e covariâncias (S_i) de cada genótipo das variáveis padronizadas. Os estimadores apropriados para o vetor de médias (2.2) e para a matriz de variâncias e covariâncias (2.3) são:

$$\bar{W}_{i\bullet} = \frac{\sum_{j=1}^q W_{ij}}{q} \quad (2.2)$$

em que

$$W_{ij} = \begin{bmatrix} W_{ij1} \\ W_{ijk} \end{bmatrix} \quad e \quad (2.3)$$

$$S_i = \frac{1}{q-1} \sum_{j=1}^q (W_{ij} - \bar{W}_{i\bullet})(W_{ij} - \bar{W}_{i\bullet})^t$$

De cada genótipo, foi obtido o índice de risco multivariado (I_i), equação (2.4), da seguinte forma:

$$I_i = \xi_{1-\alpha} \bar{W}_{i\bullet} + Z_{(1-\alpha)} \sqrt{\lambda_1} \quad (2.4)$$

em que $Z_{1-\alpha}$ é o quantil $1-\alpha$ da distribuição normal padrão; ξ_1 e λ_1 são os primeiros autovetores e autovalores de S_i , respectivamente. É interessante observar que se a explicação do primeiro autovalor for pequena, a técnica em questão não deve propiciar vantagens se for aplicada.

Os autovalores e autovetores de S_i são obtidos da maximização da forma quadrática (2.5):

$$\lambda_k = \max_{e_k} \frac{e_k^t S_i e_k}{e_k^t e_k} \quad (2.5)$$

O máximo é obtido da resolução do sistema de equações homogêneas (2.6) dado por:

$$[S_i - \lambda_k I] e_k = 0 \quad (2.6)$$

O método das potências ou o método QR pode ser usado para obter soluções para o sistema (2.6) de equações homogêneas (BOCK, 1975).

Também foi obtido um índice (Prb_i) baseado nas probabilidades normal bivariada. Para isso, o parâmetro de correlação da normal bivariada padrão foi estimado ($r_{1,2}$) para o genótipo i a partir da matriz de covariâncias S_i . Também foram estimados os valores padronizados da variável k por: $Z_{ik} = (\bar{W}_{i,k} - 100) / \sqrt{S_{i,kk}}$; em que $S_{i,kk}$ é o elemento da matriz S_i correspondente à linha e coluna k , $k = 1, 2$. Então, o índice de risco sugerido pode ser obtido por:

$$Prb_i = \frac{1}{2\pi\sqrt{1-r_{1,2}^2}} \int_{-\infty}^{Z_{i1}} \int_{-\infty}^{Z_{i2}} \exp\left[-\frac{(u^2 - 2ruv + v^2)}{2(1-r_{1,2}^2)}\right] dv du \quad (2.7)$$

em que u e v são duas variáveis normais bivariadas; $r_{1,2}$, o estimador do coeficiente de correlação entre u e v ; e Z_{i1} e Z_{i2} , dois valores padronizados de interesse da normal, que definem a região da distribuição bivariada que será integrada.

Numa última etapa, foi apresentado um exemplo ilustrativo dos resultados das análises utilizando as técnicas anteriormente descritas, usando, para isso, dados de oito experimentos realizados na Região Nordeste com nove híbridos simples de Melão. Foram consideradas nas análises as variáveis produção comercial - Prod (kg/ha) e peso de frutos - PT (kg). Para fins de comparação, foram realizadas as análises univariadas pelos métodos de Lin e Binns (1988), Annicchiarico (1992) e Annicchiarico et al. (1995).

Exemplo ilustrativo

Na Tabela 1 apresentam-se os resultados para os nove híbridos simples de melão tanto para o índice de Annicchiarico bivariado, quanto para o índice da probabilidade bivariada. Para ilustrar os cálculos, utilizou-se o híbrido 1 e calcularam-se o vetor de médias (2.2) e a matriz de covariância amostral (2.3) dos valores padronizados (2.1). Os resultados obtidos foram:

$$\bar{W}_{1\bullet} = \frac{\sum_{j=1}^q W_{ij}}{q} = \frac{\begin{bmatrix} 66,78 \\ 87,80 \end{bmatrix} + \dots + \begin{bmatrix} 103,44 \\ 98,80 \end{bmatrix}}{8} = \begin{bmatrix} 87,40 \\ 87,53 \end{bmatrix}$$

$$e S_1 = \begin{bmatrix} 146,30 & 36,01 \\ 36,01 & 42,50 \end{bmatrix}$$

Utilizando a expressão (2.6), obtiveram-se o maior autovalor e o seu correspondente autovetor, dados por:

$$\hat{\lambda}_1 = 157,6 \text{ e } \hat{e}_1 = \begin{bmatrix} 0,9544 \\ 0,2987 \end{bmatrix}$$

Portanto, o índice de risco desse híbrido, para $\alpha = 0,25$, é:

$$I_1 = \hat{e}_1^t \bar{W}_i + Z_{(1-\alpha)} \sqrt{\hat{\lambda}_1} = [0,9544 \quad 0,2987] \begin{bmatrix} 87,40 \\ 87,53 \end{bmatrix} + 0,674489 \sqrt{157,57} = 118,03$$

Os demais valores da Tabela 1 foram obtidos de forma similar. Para o índice bivariado, o processo é mais complexo: inicialmente, é substituído na expressão (2.7) o valor $r_{1,2} = 0,4556$. O valor do índice Prb_i é computado aplicando-se métodos de quadratura na expressão resultante, após especificar os valores de Z_1 e Z_2 , que podem ser, por exemplo, iguais a 1,96. Também se podem utilizar programas como o SAS[®], SAS INSTITUTE (1995) e aplicar a função "PROBBNRM($Z_1, Z_2, r_{1,2}$)", que retorna diretamente o valor da integral dupla (2.7).

Verifica-se na Tabela 1 que o híbrido de melhor desempenho foi aquele que apresentou o maior índice, ou seja, o híbrido simples número 6. Esse híbrido também foi o que apresentou maior probabilidade bivariada. Para o segundo melhor híbrido, híbrido 4, pelo índice de Annicchiarico bivariado, também se observou a segunda maior probabilidade bivariada. Não houve concordância, no entanto, para a terceira classificação. O índice de Annicchiarico bivariado apontou como terceiro melhor o híbrido de número 9, o qual obteve a quarta classificação no índice da probabilidade bivariada. O pior híbrido pelo método de Annicchiarico bivariado foi o número 5, enquanto no índice de probabilidade bivariada, foi o número 1. O híbrido número 8 recebeu a quarta classificação pelo índice de Annicchiarico bivariado e a terceira pelo índice da probabilidade bivariada. Convém salientar que os híbridos associados aos maiores valores dos índices de Annicchiarico bivariado tiveram, em geral, uma menor porcentagem da explicação do primeiro autovalor. Todavia, a explicação do primeiro componente foi relativamente satisfatória, sendo superiores a 68%, para todos os híbridos simples.

Finalmente, deve ser salientado que os dois índices tiveram uma concordância relativamente alta e significativa ($P < 0,0475$), se for observada a estimativa do coeficiente de correlação de Pearson (0,6718), ou melhor ainda, se for considerada a correlação de Spearman, a qual foi de 0,7333 ($P < 0,0246$).

Na Tabela 2 estão apresentadas as médias das variáveis Prod e PT de todos os híbridos. Verificou-se que os híbridos mais produtivos foram nessa ordem de importância o 6, 4 e 9 e os que apresentaram maiores peso médio de fruto foram 6, 4 e 2. Os híbridos mais produtivos foram aqueles que apresentaram maiores valores para os índices bivariados. Também houve concordância desses índices com os híbridos menos produtivos e com frutos mais leves, que são os híbridos 1 e 7. O híbrido mais produtivo (6) apresentou um mesmo valor padronizado de média em relação ao híbrido 4. No entanto, ele pode ser considerado mais estável por possuir menor variância, ou seja, variância de 62,6 contra 104,7. Em geral, as estimativas de correlações entre Prod e PT ao longo dos ambientes para todos os híbridos foram positivas e de valores entre 0,3 e 0,5. Exceções foram encontradas para os híbridos 2 e 5. Essas estimativas negativas de correlação são prejudiciais ao método, uma vez que altos valores de uma variável podem estar associados a pequenos valores de outra ao longo dos ambientes.

Na Tabela 3 encontram-se os índices de riscos univariados para os nove híbridos e para as duas variáveis sob estudo. Também são apresentados os valores de P_i de Lin e Binns (1988), bem como os desvios de inte-

ração de P_i , expressos em porcentagem de componentes de interação. Para Prod, novamente destacaram-se os híbridos 6 e 4, com valores superiores a 100%. Para a variável PT, resultados similares ao da variável Prod foram consistentemente observados. De maneira geral, os três piores híbridos, considerando simultaneamente as duas variáveis, foram o 1, 7 e 2. O híbrido 5, destacado como o pior no índice de risco multivariado, esteve entre os híbridos de performance intermediária, considerando os índices univariados. Em ambas as variáveis, o híbrido 5 apresentou valores menores do que 100% para o índice I, mesmo considerando o valor de 25% para a significância nominal. Esse valor de 25% é recomendado por Annicchiarico (1992) e Annicchiarico et al. (1995).

Com a decomposição de P_i , verificou-se que as maiores contribuições para a interação foram dos híbridos 2, 7 e 3 para Prod e dos híbridos 9, 7 e 1 para PT. Surpreendentemente para PT, o híbrido 4 apresentou a quarta maior contribuição de P_i para a interação, uma vez que o valor de P_4 foi considerado não significativamente diferente do máximo. Com relação ao valor de P_i , pode-se afirmar que os híbridos 4, 6 e 9 não apresentaram valores médios significativamente diferentes do máximo ao longo dos ambientes para Prod, sendo um forte indício de serem materiais adaptados e estáveis. Para a variável PT, os híbridos 6 e 4 não apresentaram P_i significativamente diferentes de zero. Todos os demais genótipos apresentaram valores de P_i diferentes de zero, indicando serem genótipos não-estáveis e não-adaptados.

TABELA 1 – Índices de risco de Annicchiarico bivariado (I_i) com suas respectivas explicações da variação retida pelo primeiro componente principal (PEXPL_i) e índice derivado da probabilidade bivariada (Prb_i) para o i-ésimo híbrido simples de melão.

| Híbridos simples (i) | I_i | PEXPL _i | Prb _i |
|------------------------|--------|--------------------|------------------|
| 1 | 118,03 | 83,39 | 0,01 |
| 2 | 88,16 | 91,68 | 0,21 |
| 3 | 114,66 | 93,21 | 0,22 |
| 4 | 153,76 | 71,40 | 0,79 |
| 5 | 76,65 | 69,69 | 0,14 |
| 6 | 162,23 | 68,07 | 0,87 |
| 7 | 135,51 | 80,27 | 0,10 |
| 8 | 141,08 | 73,71 | 0,32 |
| 9 | 141,27 | 74,56 | 0,27 |

TABELA 2 – Valores médios dos dados originais $\bar{Y}_{i \cdot k}$ e padronizados $\bar{W}_{i \cdot k}$, variância $S_{kk}^{(i)}$ e correlação de Pearson $r_{1,2}^{(i)}$ dos dados padronizados para o i-ésimo híbrido ($I = 1, 2, \dots, 9$) e para k-ésima variável, qual seja, produção comercial em kg/ha (Prod, $k = 1$) e peso do fruto em kg (PT, $k = 2$).

| Híbridos Simples | PROD | | | PT | | | $r_{1,2}^{(i)}$ |
|---------------------|-----------------------|-----------------------|----------------|-----------------------|-----------------------|----------------|-----------------|
| | $\bar{Y}_{i \cdot 1}$ | $\bar{W}_{i \cdot 1}$ | $S_{11}^{(i)}$ | $\bar{Y}_{i \cdot 2}$ | $\bar{W}_{i \cdot 2}$ | $S_{22}^{(i)}$ | |
| 1 | 34952,94 | 87,4 | 146,3 | 1,20 | 87,5 | 42,7 | 0,4556 |
| 2 | 37032,06 | 94,6 | 448,3 | 1,42 | 103,9 | 56,5 | -0,4824 |
| 3 | 40828,13 | 103,7 | 147,0 | 1,35 | 98,5 | 10,8 | 0,0995 |
| 4 | 43269,16 | 111,2 | 104,7 | 1,50 | 109,5 | 58,7 | 0,3360 |
| 5 | 38811,28 | 98,6 | 75,4 | 1,35 | 98,3 | 36,8 | -0,2044 |
| 6 | 43558,72 | 111,2 | 62,6 | 1,56 | 114,2 | 87,3 | 0,3258 |
| 7 | 36046,56 | 90,7 | 168,2 | 1,24 | 90,4 | 103,7 | 0,5733 |
| 8 | 39351,28 | 99,5 | 82,6 | 1,38 | 100,7 | 46,8 | 0,4005 |
| 9 | 41212,94 | 103,1 | 180,9 | 1,33 | 96,9 | 86,1 | 0,3627 |

TABELA 3 – Índices de riscos (I_i) e valores de P_i com as respectivas contribuições para a interação genótipos x ambientes (GxA), para os nove híbridos e para as variáveis produção comercial em kg/ha (Prod) e peso total de frutos em kg (PT).

| Híbridos Simples | Prod | | | PT | | |
|---------------------|--------|-----------------|-----------------------------|--------|---------|-----------------------------|
| | I_i | P_i | % contribuição p/ GxA | I_i | P_i | % contribuição p/ GxA |
| 1 | 79,24 | 352641661,671* | 10,21 | 83,12 | 0,375* | 15,67 |
| 2 | 80,32 | 364766231,797* | 24,25 | 98,87 | 0,093* | 7,13 |
| 3 | 95,48 | 181145247,734* | 15,31 | 96,32 | 0,141* | 2,41 |
| 4 | 104,25 | 54199342,469ns | 4,12 | 104,29 | 0,066ns | 15,45 |
| 5 | 92,76 | 181857203,219* | 7,67 | 94,21 | 0,159* | 9,74 |
| 6 | 105,87 | 68133733,688ns | 6,58 | 107,90 | 0,016ns | 3,73 |
| 7 | 81,99 | 348454155,016* | 16,43 | 83,53 | 0,313* | 16,03 |
| 8 | 93,39 | 171131635,594* | 8,48 | 96,10 | 0,144* | 13,04 |
| 9 | 94,04 | 111758835,578ns | 6,94 | 90,67 | 0,210* | 16,79 |
| Cutoff | | 138311346,099 | | | 0,067 | |

*, ns: significativo e não-significativo, pelo teste F a 5% respectivamente

CONCLUSÕES

Foi obtida a derivação teórica da extensão bivariada do índice de risco de Annicchiarico com sucesso e foi proposto um segundo índice de risco baseado na probabilidade bivariada; os dois índices apresentaram grande concordância, considerando um exemplo ilustrativo com genótipos de melões.

REFERÊNCIAS BIBLIOGRÁFICAS

- ACCIARESI, H. A.; CHIDICHIMO, H. O. Genotype-environment interaction in Avena sativa L-employing AMMI and factorial correspondence models. **Pesquisa Agropecuária Brasileira**, Brasília, v. 34, n. 10, p. 1823-1830, out. 1999.
- ANNICCHIARICO, P. Cultivar adaptation and recommendation from alfalfa trials in Northern Italy. **Journal Genetics and Breeding**, v. 46, n. 1, p. 269-278, Mar. 1992.
- ANNICCHIARICO, P.; BERTOLINI, M.; MAZZINELLI, G. Analysis of genotype-environment interactions for maize hybrids in Italy. **Journal of Genetics and Breeding**, Rome, v. 49, p. 61-68, 1995.
- BECKER, H. C.; LEON, J. Stability analysis in plant breeding. **Plant Breeding**, New York, v. 101, n. 1, p. 1-23, 1988.
- BOCK, R. D. **Multivariate statistical methods in behavioral research**. New York: McGraw Hill, 1975.
- CROSSA, J. Statistical analysis of multilocations trials. **Advances in Agronomy**, New York, v. 44, p. 55-85, 1990.
- GAUCH JUNIOR, H. G.; ZOBEL, R. W. Predictive and postdictive success of statistical analysis of yield trials. **Theoretical and Applied Genetics**, Berlin, v. 76, n. 1, p. 1-10, 1988.
- JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 4. ed. New Jersey: Prentice-Hall, 1998. 816 p.
- LIN, C. S.; BINNS, M. R. A superiority measure of cultivar performance for cultivar x location data. **Canadian Journal of Plant Science**, Ottawa, n. 68, p. 193-198, Jan. 1988.
- SAS INSTITUTE. **SAS language and procedures: usage: version 6**. Cary, 1995. 373 p.
- YAU, S. K. Regression and AMMI analysis of genotype x environmental interactions: an empirical comparison. **Agronomy Journal**, Madison, v. 87, n. 1, p. 121-126, Jan./Feb. 1995.